EDRN Public Portal Upgrade

Including the EDRN Knowledge Environment 2008

NASA Jet Propulsion Laboratory 4800 Oak Grove Drive Pasadena CA 91109-8001 http://www.jpl.nasa.gov/

Table of Contents

Introduction

The EDRN Public Portal has been running for over a year providing a public face to our efforts and goals. We are now ready to upgrade the portal with new applications to provide functions and features to EDRN members in managing scientific data, tracking biomarker status, and so forth. These applications include:

- EDRN-specific content objects
- EDRN Knowledge Environment (EKE)
- LDAP-based authentication and authorization

Requirements

The EDRN Public Portal must satisfy a number of requirements as it serves as the public face and image of the Early Detection Research Network. While these requirements are mainly qualitative, they reflect the vision of the staff of EDRN in terms of maintaining that public face, as well as providing advanced services for EDRN members.

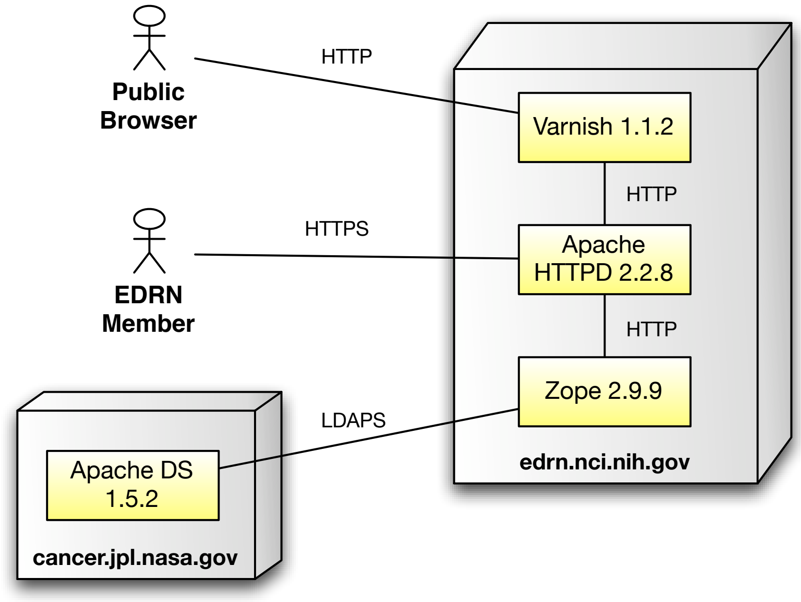
The requirements are:

- Serve public visitors using a web browser and provide timely information about EDRN
- Service EDRN members securely with encrypted communications and provide sensitive information about EDRN
- Serve as a platform for applications for EDRN members
- Adhere to the latest recommendations for web server security
- Satisfy NCI security scans using IBM Rational AppScan with zero critical issues, zero high issues, two medium issues, and twelve low issues.
- Enable automatic updating of site content
- Enable EDRN staff with appropriate privileges to update site content

Additional requirements may be found in the original specification documents for the EDRN Public Portal. These requirements should be considered advisory, and may be modified as needed in order to facilitate fast delivery and operational mode of the EDRN Public Portal upgrade.

System Architecture

The following diagram depicts the architecture of the system to run the EDRN Public Portal.



The host edrn.nci.nih.gov is maintained by Terrapin Systems for NCI. NASA Jet Propulsion Laboratory maintains the host cancer.jpl.nasa.gov. Members of the general public view the portal through regular HTTP over port 80 through the Varnish cache front end. Members of EDRN may log in securely to the portal to access additional features over HTTPS through port 443.

Apache HTTPD provides URL rewriting services as well as HTTPS in order to serve both public and members-only access. Apache will listen to port 443 publicly for HTTPS and internally answer HTTP queries from the Varnish cache on port 8100 (or another convenient port).

The Zope application server runs the EDRN Public Portal Application and includes an integrated database. It will listen internally on port 8101 (or other convenient port) to accept requests for page compositing and other data from Apache HTTPD. When EDRN members log in, it will open LDAPS connections to cancer.jpl.nasa.gov to authenticate users against the EDRN Directory Service, maintained by NASA.

Firewall Requirements

The following table lists the required firewall exceptions that must be in place for the hosts edrn.nci.nih.gov and cancer.jpl.nasa.gov in order to support operation of the EDRN Public Portal:

Host	Port	Direction	Protocol	Purpose
edrn.nci.nih.gov	80/TCP	Incoming	HTTP	Web browsers
edrn.nci.nih.gov	443/TCP	Incoming	HTTPS	Authenticated browsers
edrn.nci.nih.gov	Any/ TCP	Outgoing	LDAPS	Authentication requests to port 636 on cancer.jpl.nasa.gov
cancer.jpl.nasa.gov	636/TCP	Incoming	LDAPS	Authentication requests from edrn.nci.nih.gov

Installing the Components

The EDRN Public Portal is composed of several software components from a variety of sources, mainly open source software and software developed by the NASA Jet Propulsion Laboratory. Together, these components serve the needs of the portal's audience and provide an upgrade path for future enhancements.

Installation at NCI

Terrapin Systems will need to install (or upgrade) the software components on the machine that will serve as edrn.nci.nih.gov, as summarized in the following table:

Component	Version	Source	Purpose
Varnish	1.1.2	http://varnish-cache.org/	High performance HTTP accelerator
Apache HTTPD	2.2.8	http://httpd.apache.org/	Web server
Python	2.4.5	http://python.org/	Programming language for application server
PIL	1.1.6	http://www.pythonware.com/products/pil/	Imaging library for Python, required by Plone
elementtree	1.2.6	http://effbot.org/downloads/	XML toolkit
RDFLib	2.3.3	http://rdflib.net/	Resource Description Format API

Component	Version	Source	Purpose
Python-LDAP	2.3.4	http://python-ldap.sourceforge.net/	LDAP client API
Zope	2.9.9	http://zope.org/	Application server
Plone	2.5.5	http://plone.org/	Content management system
PloneLDAP	1.0	http://plone.org/products/ploneldap	LDAP authentication/authorization for Plone
CMFSin	SVN 56454	http://svn.plone.org/svn/collective/CMFSin/trunk	RSS aggregator; please use revision 56454 from Subversion repository
EDRNPortalSkin	1.1.1	http://agility.jpl.nasa.gov/products/edrnportalskin	Look and feel for EDRN
EDRNContent	1.0.0	http://agility.jpl.nasa.gov/products/edrncontent	EDRN-specific content objects
EKE	1.0.1	http://agility.jpl.nasa.gov/products/eke	EDRN Knowledge Environment

Installation at NASA Jet Propulsion Laboratory

Software components at NASA are already deployed on the host cancer.jpl.nasa.gov and consist of one package:

Component	Version	Source	Purpose
Apache DS	1.5.1	http://directory.apache.org/	EDRN Directory Service via LDAPS

Note that since the EDRN Directory Service uses LDAPS, the TLS certificate will need to be installed on edrn.nci.nih.gov so that the application server can successfully communicate with the authentication server. This certificate will be made available in PEM format by NASA/JPL on request.

Configuring the Components

Integrating an application server, HTTPS, and caching HTTP acceleration requires some fairly precise configuration. In order to simplify such installation, NASA Jet Propulsion Laboratory has already determined a functioning set of configuration for the server components to be installed on `edrn.nci.nih.gov`.

Configuring the HTTP Accelerator

Varnish uses a combination of both a configuration file and command-line arguments in order to function. Its purpose is to request complete HTTP objects from an upstream provider (Apache HTTPD in this case), cache them for future use, and then deliver them to a hungry web audience.

Varnish uses instructions in a configuration file (which it compiles at startup into executable code). In addition to the configuration file, Varnish requires some command-line arguments at startup in order to tell it where the cache file is on disk, what port number to listen to, what kinds of HTTP requests to serve, and what management port to use.

A Linux-compatible "init" script that sets the correct command-line parameters, as well as the Varnish configuration file, are available at the following address:

http://tinyurl.com/4z9pme

For each file, make the simple substitutions listed on the above web page.

Configuring the Web Server

Sitting in between the HTTP accelerator and the application server is the Apache HTTPD web server. Its job is to rewrite incoming page URLs to the appropriate requests for the Zope application server and the Plone content management system. It also redirects users to HTTPS connections when detecting requests for the Plone login page as well as if the user is already logged in, requiring her to maintain HTTPS connections. Lastly, it listens on HTTPS port 443, as neither Varnish nor Zope support SSL/TLS.

This section details the configuration necessary for Apache HTTPD.

Activating Encryption

In order to protect passwords and well as any sensitive data, the Apache web server must support HTTPS using SSL/TLS. Terrapin Systems on behalf of NCI will purchase an appropriate encryption certificate from a well-known certificate authority such as Thawte or VeriSign.

Once the certificate is signed by the authority, any password on the certificate file should be removed so that the web server may start without human intervention.

Details of this process is left to Terrapin Systems and NCI as they are dependent on the certificate authority used.

Adjusting the Configuration File

The configuration file for Apache HTTPD, httpd.conf, should be adjusted with large number of settings. Rather than include them here, a sample configuration file is available at the address

http://agility.jpl.nasa.gov/documents/edrn-public-portal

under the link "Apache HTTPD Configuration".

Configuring the Application Server

The EDRN Public Portal leverages the Plone content management system in order to provide a logical layout, indexing, look-and-feel, search, storage, and other content features. Plone runs under the Zope application server, an open-source and highly secure system for building web applications.

To configure Zope, you'll need to adjust some settings in its configuration file and then install the add-on products that comprise the EDRN Public Portal. (Note that Zope itself requires Python 2.4, and the add-on products require that the Python installation include the Python Imaging Library (PIL), the Python-LDAP API, and the Resource Description Format Library (RDFLib). Ensure when installing these packages that they are installed into the same Python installation that Zope itself is using.)

Adjusting Zope's Configuration File

Configuring Zope merely requires setting of the following properties in the "zope.conf" configuration file:

- effective-user: should be set to "nobody" or another sandbox user
- datetime-format: international
- rest-input-encoding: utf-8
- rest-output-encoding: utf-8
- access logger: comment-out completely
- default-zpublisher-encoding: utf-8
- http-server address: 8101 (or other convenient private port; ensure that Apache HTTPD is configured to pass requests to this port number)

Installing Add-on Products

Add-on products for Zope consist or TAR or ZIP archives that are extracted into the Zope instance's "Products" directory.

The add-on products necessary for the EDRN Public Portal include

- Plone

- PloneLDAP
- CMFSin
- EDRNPortalSkin
- EDRNContent
- EKE

See the table in the section “Installation at NCI” for more information on each of these.

Once the add-on packages are installed, restart the Zope application server. Then use the Zope management interface to add a “Plone Site” object to the root of the Zope object tree named “edrn”. Visit the “edrn” object’s plone_control_panel and use the Add/Remove products form to add CMFSin, EDRNPortalSkin, EDRNContent, and EKE.

Configuring LDAP Authentication

To enable LDAP authentication, use the Zope management interface and perform the following steps:

1. Navigate to the `acl_users` object.
2. From the add menu, add an LDAP Multi Plugin. Fill out the fields as follows:
   - ID: `ednrdir`
   - Title: `EDRN Directory Service`
   - LDAP Server: `cancer.jpl.nasa.gov`
   - Use SSL: Yes
   - Read Only: Yes
   - Login name attribute: `uid`
   - User ID attribute: `cn`
   - RDN attribute: `uid`
   - Users base RDN: `dc=edrn,dc=jpl,dc=nasa,dc=gov`
   - Scope for users base: ONELEVEL
   - Group storage: Groups not stored
   - Group base RDN: `dc=edrn,dc=jpl,dc=nasa,dc=gov`
   - Scope for group base: ONELEVEL

- Manager DN: (leave blank)
- Manager's password: (leave blank)
- User password encryption: clear
- Default user roles: `Anonymous`

3. Navigate to the new `edrndir` object.
4. Click the Contents tab.
5. Navigate to the embedded `acl_users` object.
6. For the User Object classes, enter: `top,person,organizationalPerson,inetOrgPerson,edrnPerson`
7. Click Apply Changes
8. Click the LDAP Schema tab
9. Check the box next to `cn` and click Delete.
10. Add an LDAP Schema Item with LDAP Attribute Name `cn`, Friendly Name `Friendly Name`, Multi Valued no, and Map to Name `fullname`, then click Add.
11. Add another LDAP Schema Item with LDAP Attribute Name `mail`, Friendly Name `Email Address`, Multi Valued no, and Map to Name `email`, then click Add.
12. Add yet another LDAP Schema Item with LDAP Attribute Name `description`, Friendly Name `Description`, Multi Valued no, and Map to Name `description`, then click Add.
13. Navigate back to the Plone object and then to the `acl_users` object. From there, navigate to the plugins object.
14. For each type of plugin (Anonymoususerfactory Plugins, Authentication Plugins, Challenge Plugins, etc.), navigate to the plugin. If `edrndir` appears in the list of available plugins, select it and move it to the list of active plugins. Also, use the arrows by the active plugins to move it to the top of the list. Repeat this for all 24 plugins. (Note that for most of the plugins the `edrndir` is not available anyway, simply because it does not provide that particular authentication/authorization service.)

At this point, LDAP authentication should be active. To test it, use the test user account set up in the EDRN Directory Service and log into the Plone site:

- User name: `edrndemo`

- Password: `edrndemo`

Importing Old Content

Content from the previous EDRN Public Portal will not be required. All content will be of a new form managed by EDRN and NASA staff.

Contact Information

This summary was written by Sean Kelly of the EDRN Informatics Center. The EDRN Informatics Center is run by NASA's Jet Propulsion Laboratory which is operated by the California Institute of Technology.

EDRN Informatics Center
NASA Jet Propulsion Laboratory
Daniel J Crichton, Primary Investigator
MS 169-315
4800 Oak Grove Drive
Pasadena, California 91109-8099